PROJECT REPORT

Indexing and Compression

Bhattacharya, Abhirup

Description of the Code:

1. A block size of 20000 has been used to create the individual block files with the postings, which has later been merged into a single file. This has been done separately for each of 2 versions of the index to get the final Index_Version1.uncompress and Index_Version2.uncompress after merging.

2. Each of the document in the cranfield collection has been opened and tokenized (using the tokenizer implemented in Assignment 1). After tokenizing, the NLTK version of the porter stemmer has been used for stemming the tokenized term and Wordnet Lemmatizer has been used for lemmatization.

3. Once number of terms in memory crosses 20000, the current document being processed is completed and the block index file is written to the disk. These files are of the form "lemma_posting_*fileno*" and stem_posting_*fileno*".

4. I have implemented the single-pass in-memory indexing(SPIMI) to create the individual block files of the indexes and then merged into the final uncompressed postings.

5. For the first step while merging, we will read the first line from each of the block files and create a dictionary with the file handle and the corresponding line (we will call it the file_dict). The terms are then compared with each other, and the lowest lexicographical term and its corresponding entry are written to the final merged posting file. Then the next line from the file from which the last line was written is read and added to the file_dict with the corresponding file_handle. And the process is repeated till any file reaches the end. Once a file is read completely, it is removed from the file_dict dictionary. The process stops when all the files are removed from the file_dict.

6. Also, while creating these final uncompressed posting files, an uncompressed dictionary has been created in memory using a hashmap, with the term as the key and the df and posting pointer as the corresponding value. The posting pointer points to the line on the Index_Version1.uncompress and Index_Version2.uncompress files, which can be used to fetch the corresponding postings for any given term without having to parse the entire file. Thus, with the hashmap implementation, we can fetch the posting of any file in constant time complexity O(1).

7. A sample entry in the final posting files is shown below:

    ablating:2, (553, 3, 3, 171), (1100, 1, 4, 114)

    ablat:13, (1097, 6, 9, 225), (553, 6, 6, 171), (1065, 1, 3, 114), (274, 2, 5, 228), (1279, 1, 4, 158), (1096, 1, 2, 61), (1241, 2, 4, 152), (1226, 2, 8, 245), (1099, 3, 3, 64), (1101, 2, 3, 88), (1098, 2, 5, 143), (82, 1, 10, 284), (1100, 3, 4, 114)

The entry are of the form:

Term:df,(doc_id,tf,max_tf,doclen)

Where **df :** Number of documents in which the term has occurred

Doc_id : The document ID in which the term has occurred

Tf : Number of times the term has occurred in the above document

Max_tf : The largest TF of the terms in the current document

Doclen : The length of the current document

The terms marked in red form a single tuple, and this tuple can be repeated the number of time same as the df (one entry for each document)

Compression:

8. A function has been created to compress each of the 2 versions of the index, taking argument as the filetype(1 for version1 and 2 for version2).

9. In case of Version1, the dictionary has been compressed using the blocked compression method with k=4. A part of the term string is shown below for version1:

   2ab11abbreviated7ability13ablatedlength8ablating8ablation8ablative4able5about5above14 abovementioned6abrupt8abruptly7absence6absent8absolute8absorbed9absorbing10absorp tion8abstract10abundantly8academic11accelerated11accelerates12accelerating

   And the corresponding dictionary entries:

   (1, 0, 0)
   (1, 24)
   (3, 57)
   (1, 121)
   (2, 157, 39)
   (13, 205)
   (1, 458)
   (9, 488)
   (171, 658, 71)
   (68, 3799)
   (1, 5045)
   (3, 5079)
   (3, 5145, 106)
   (9, 5211)
   (1, 5390)
   (11, 5418)
   (8, 5633, 139)
   (1, 5794)
   (6, 5826)
   (2, 5950)
   (1, 5997, 179)
   (1, 6029)
   (5, 6060)
   (1, 6162)
   (4, 6194, 226)
   (18, 6281)
   (2, 6623)

(1, 6671)

The entry is of the form document frequency, posting pointer (pointing to the line in the compressed posting file where the term occurs), term pointer (pointing to the part of the term string where the term occurs. The term pointer is only stored for every 4th word of the dictionary, and every other word can be reached by following the length of the words mentioned in the term string. The dictionary has been stored in disk and also in the memory for faster access.

10. The posting has been compressed using gamma encoding, considering only the gaps in the doc_ids, and discarding the tf, max_tf and the doclen. A sample entry of the compressed posting file is shown below:

ablating:2,['111111111110000101001', '11111111110000100011']

The posting file is named as Index_Version1.compress.

11. For Version2 index, the dictionary has been compressed using front encoding. A sample of the term is shown as follows:

ab*out2\$ov7\$ovement4\$rupt

I have used the \$ symbol in place of the diamond symbol as shown in class notes.

12. The dictionary entries are similar to that for version1. A snippet of the dictionary file is shown below:

(1, 0, 0)
(1, 24)
(3, 53)
(10, 114)
(13, 303)
(1, 553, 37)
(171, 589)
(68, 3730)
(1, 4975)
(3, 5004, 63)
(3, 5070)
(9, 5136)
(1, 5314)

The format is same as that in case of Version1 dictionary. The dictionary has been stored in disk and also in the memory for faster access.

13. The posting has been compressed using delta encoding, considering only the gaps in the doc_ids, and discarding the tf, max_tf and the doclen. A sample entry of the compressed posting file is shown below:

ablat:13,['1110011000100101', '1110010000100000', '111001000000000', '1110010100010111', '1110010111101101', '1110000011011', '1110000010001', '11000111', '11011111111', '1000', '1001', '1110010111111000', '1110010111111010']

The posting file is named as Index_Version2.compress.

Finding Postings for the terms in problem statement:

14. To find the postings for each of the given words, we can use the uncompressed dictionary with the terms to find the posting pointer, which in turn can be used to get the posting from the posting file stored in the disk. The TF is thus obtained from the term postings and the sum of all the TFs of all the documents are added to find the total TF of the given term. The total posting length of the postings in byte is also computed from the obtained posting list.

15. Same logic is used to find the posting list for the term NASA. The document frequency, TF and Postings for the first 3 entries of the posting list.

16. To find the term(s) with the largest and the smallest DF, the dictionary is iterated to find the document frequencies and the terms with the highest and smallest frequencies are appended to a list and printed. The result is shown in the next part of the report. The same is done for the dictionaries of both the versions.

17. To find the document with the largest doc len, the posting file is iterated through each term to find the doc lens. The documents already searched are added to a set, which is checked each time the len is being read. If the document is already present in the set, we don't read the doc len.

Statistics for the program and index files:

1. Time taken for building indices:

  1a. Time taken for building Index_Version1.uncompress : 5.973879337310791 seconds

  1b. Time taken for building Index_Version2.uncompress : 6.369318723678589 seconds

  1c. Time taken for building Index_Version1.compress   : 0.6033968925476074 seconds

  1d. Time taken for building Index_Version2.compress   : 0.5904693603515625 seconds

2. Size of Index_Version1.uncompress 1757148 Bytes

3. Size of Index_Version2.uncompress 1685572 Bytes

4. Size of Index_Version1.compress 2051103 Bytes

5. Size of Index_Version2.compress 1743903 Bytes

6. Number of Postings in each Version:

 6a. Number of Postings in Version1 Index: 6299

 6b. Number of Postings in Version2 Index: 4498

7. Postings for given list of words:

TERM : reynolds

 Df : 192

 Total TF : 358

 Inverted Posting Lentgh (in Bytes) : 1800

TERM : nasa

 Df : 7

 Total TF : 6

 Inverted Posting Lentgh (in Bytes) : 160

TERM : prandtl

 Df : 63

 Total TF : 79

 Inverted Posting Lentgh (in Bytes) : 576

TERM : flow

Df : 678

Total TF : 1696

Inverted Posting Lentgh (in Bytes) : 5888

TERM : pressure

Df : 529

Total TF : 1227

Inverted Posting Lentgh (in Bytes) : 4576

TERM : boundary

Df : 457

Total TF : 997

Inverted Posting Lentgh (in Bytes) : 4024

TERM : shock

Df : 235

Total TF : 658

Inverted Posting Lentgh (in Bytes) : 2072

8. DF,TF and Postings for NASA

 DF for NASA : 7

 Entry 1 values :

 TF   : 1

 Doclen : 103

 MaxTF : 4

 Entry 2 values :

 TF   : 1

 Doclen : 100

 MaxTF : 4

Entry 3 values :

TF   : 1

Doclen : 195

MaxTF : 6

9. Dictionary Term with Largest and Lowest DFs in Index Version 1

VERSION 1 MAX and MIN DFs :

Max DF Terms(s) : ['to']

Min DF Terms(s) : ['ab', 'abbreviated', 'ablatedlength', 'ablative', 'abovementioned', 'absent', 'absorbing', 'abundantly', 'academic', 'accelerates', 'accelerometer', 'accentuated', 'acceptability', 'acceptably', 'acceptance', 'accessible', 'accidental', 'accommodate', 'accommodated', 'accompany', 'accountable', 'accounting', 'accrue', 'accumulated', 'acetate', 'achievable', 'acoustical', 'acoustically', 'acquire', 'acquired', 'acquisition', 'acr', 'acrodynamic', 'acrothermochemistry', 'acrothermoelasticity', 'actural', 'adaptable', 'additive', 'adjacentequilibrium', 'adjoining', 'adjoint', 'adjusting', 'administration', 'admit', 'admitted', 'admitting', 'admixture', 'adsorption', 'advent', 'aeolotropic', 'aerfoils', 'aerial', 'aeroballistics', 'aerocosmonautics', 'aerodynamieist', 'aerothermal', 'aerothermochemical', 'aerothermodynamic', 'aerothermoelastic', 'affinely', 'affinity', 'affirmative', 'afford', 'afforded', 'afresh', 'afterburner', 'afterflow', 'age', 'agency', 'agreeing', 'ailcron', 'aimed', 'airborne', 'aircrafttype', 'airforce', 'airframe', 'airliner', 'airload', 'airloads', 'airscrew', 'airspeed', 'akin', 'aligned', 'allen', 'alleviating', 'allmovable', 'allround', 'alluded', 'almen', 'alternately', 'alternating', 'alternatively', 'altogether', 'aluminium', 'amer', 'america', 'amongst', 'amounted', 'ample', 'amplifier', 'amplifying', 'amply', 'analogously', 'analoguous', 'analyser', 'analyticially', 'analyzable', 'analyzer', 'analyzes', 'andlanding', 'anh', 'anhedral', 'anisotropy', 'annex', 'announces', 'annual', 'annul', 'anove', 'anti', 'antielastic', 'antisymmetric', 'antisymmetrical', 'anyway', 'ao', 'apogee', 'apparentmass', 'appraisal', 'appraised', 'appraisement', 'appraoch', 'approximates', 'apropriate', 'ar', 'ardc', 'arithmetic', 'arithmetical', 'arrest', 'arrived', 'arrow', 'ary', 'ascertained', 'ascribed', 'aside', 'asimplified', 'ask', 'aspectratio', 'assembly', 'asserted', 'assessing', 'assistance', 'association', 'assortment', 'assure', 'astronautics', 'astrophysics', 'atlarge', 'attach', 'attacked', 'attacking', 'attaindd', 'attaining', 'attainment', 'attened', 'attract', 'aural', 'auspex', 'authored', 'authority', 'availability', 'avaluated', 'avove', 'avro', 'axialthrust', 'backed', 'backing', 'backup', 'backward', 'bagley', 'bakanov', 'balanced', 'balsa', 'ban', 'bang', 'bare', 'barrel', 'basetype', 'basin', 'basing', 'batchelor', 'batdorf', 'bc', 'beat', 'bed', 'bedford', 'begs', 'behaved', 'believe', 'belleville', 'belltype', 'belong', 'belt', 'bendingtorsion', 'bent', 'ber', 'bergdolt', 'berger', 'bernoulli', 'berthelot', 'bertram', 'beskin', 'beta', 'beveled', 'beyween', 'bifurcated', 'bifurcation', 'billowing', 'bimetallic', 'bimolecular', 'birnbaum', 'bisector', 'bladed', 'bleich', 'blended', 'blocking', 'blown', 'blowout', 'blue', 'bluish', 'blunttrailing', 'board', 'boat', 'boattails', 'boiloff', 'bolshakov', 'bomb', 'bonus', 'booming', 'boost', 'bore', 'bornel', 'boundwell', 'bracket', 'branchline', 'braunschweig', 'breaking', 'breguet', 'bridge', 'bridging', 'brightness', 'broadened', 'broadening', 'broke', 'brooklyn', 'brown', 'budiansky', 'builtup', 'bulent', 'bulging', 'bulkhead', 'bullet', 'bump', 'buoyancy', 'buried', 'butler', 'buzz', 'bypass', 'cabin', 'calcualted', 'calcualtions', 'caliber', 'calibrate', 'cambridge', 'camera', 'canard', 'cancel', 'cancellation', 'canted', 'capillary', 'caplan', 'capsule', 'caravelle', 'carborundum', 'caret', 'carrier', 'carter', 'casing', 'cast', 'castigliano', 'castolite', 'catalytic', 'catalytically', 'catered', 'cathode', 'cavitating', 'cellulose', 'centaur', 'centerof', 'centigrade', 'centreline', 'centripetal', 'centroid', 'certainty', 'cf', 'chain', 'chap', 'chapter', 'characteristies', 'characterize', 'characterizes', 'charging', 'charter', 'chase', 'chem', 'chemisorption', 'chernyi', 'chesky', 'chester', 'chien', 'child', 'china', 'choke', 'choleski', 'chu', 'circling', 'circuit', 'circuitry', 'circulated', 'circulating', 'circulatory', 'circumlunar', 'claim', 'claimed', 'clamping', 'clamshell', 'clarified', 'clarify', 'clarifying', 'classed', 'classic', 'classically', 'classificaiton', 'classified', 'clausing', 'clay', 'clearance', 'cleared', 'climinated', 'closeness', 'closest', 'closure', 'coalescence', 'coarse', 'coasting', 'coating', 'coefstant', 'coherent',

'colder', 'collaboration', 'collaborator', 'collapsing', 'collectively', 'collocation', 'colour', 'columnmaterial', 'combine', 'combustible', 'combustionheated', 'combution', 'command', 'commence', 'commercially', 'communication', 'companion', 'company', 'compatible', 'compensated', 'compete', 'competition', 'compilation', 'complimentary', 'compound', 'compressibleflow', 'compressing', 'compressional', 'comprise', 'conceivable', 'conceivably', 'conceptual', 'concise', 'concisely', 'concocted', 'concrete', 'concur', 'concurrently', 'concurring', 'condamped', 'condensed', 'conduct', 'conductioncontrolled', 'conductionexpansion', 'conductor', 'conecylinders', 'coned', 'confidence', 'confine', 'confirming', 'conform', 'conformally', 'conformaltransformation', 'conformity', 'conforms', 'confused', 'confusing', 'conically', 'conicals', 'conidtions', 'connect', 'connecting', 'connexion', 'conservativeness', 'consierable', 'consolidate', 'const', 'constancy', 'constantdensity', 'constantly', 'constituted', 'constitution', 'constrained', 'consumption', 'contacting', 'contaminates', 'contaminating', 'contemplated', 'context', 'continually', 'continues', 'contoured', 'contract', 'contracting', 'contradiction', 'contrasting', 'contributes', 'contributing', 'convect', 'convected', 'conveniently', 'convergentdivergent', 'converse', 'convert', 'conviction', 'convolution', 'cooler', 'cope', 'copper', 'cornell', 'corollary', 'corpuscular', 'corrective', 'correctness', 'correlograms', 'cosine', 'cosmonautics', 'coulomb', 'council', 'counteracting', 'counterclockwise', 'counting', 'couple', 'coupon', 'cps', 'crabtree', 'crack', 'cracking', 'craft', 'crank', 'creased', 'creepbuckling', 'creeping', 'crest', 'crew', 'crinoline', 'criticized', 'crooked', 'crosscorrelation', 'crossing', 'crosssections', 'crushed', 'cubics', 'cumbersome', 'cure', 'curling', 'curtain', 'curving', 'cusp', 'cussed', 'cutoff', 'cutting', 'cx', 'cylinderhemisphere', 'daley', 'dalton', 'damaging', 'dampometer', 'dana', 'danger', 'dangerous', 'dashin', 'dashn', 'dashthat', 'dashthe', 'daytime', 'debatable', 'decade', 'decelerate', 'decelerated', 'decelerates', 'decide', 'decidedly', 'decision', 'decisively', 'declassified', 'declined', 'decomposed', 'decomposition', 'deducted', 'deeper', 'deepest', 'defference', 'definable', 'defines', 'deflagration', 'deformable', 'deforming', 'defy', 'degraded', 'degreec', 'degreer', 'deionization', 'deisgn', 'delft', 'delimited', 'delineated', 'demarcation', 'dence', 'denomination', 'denser', 'dent', 'depart', 'departs', 'depended', 'dependency', 'depicting', 'depletion', 'deposited', 'depression', 'derivable', 'deryagin', 'describable', 'descriptively', 'designated', 'desire', 'desk', 'detect', 'detectable', 'detection', 'deteriorate', 'deteriorates', 'determinable', 'determinant', 'determinantal', 'determinate', 'detra', 'developable', 'devise', 'devising', 'devreased', 'dewpoint', 'diagonal', 'diam', 'dicsovering', 'dictated', 'dielectric', 'differed', 'differentiate', 'differentiates', 'differentiation', 'diffuse', 'diffusely', 'diffusive', 'digit', 'dilation', 'dilute', 'dimensioning', 'dipole', 'dirac', 'directive', 'director', 'dirichlet', 'disappeared', 'discerned', 'discernible', 'discharging', 'discretefrequency', 'discussing', 'disequilibrium', 'dislocation', 'dismissed', 'disparity', 'dispersion', 'display', 'disposable', 'disregarded', 'dissertation', 'dissimilar', 'dissociate', 'dissociationrecombination', 'distant', 'distort', 'distorting', 'distributes', 'distributing', 'disturb', 'diurnal', 'diverse', 'diversified', 'divide', 'doctor', 'dodecagonal', 'dome', 'dominating', 'doppler', 'dorodnitsyn', 'dorodnitzinhowarth', 'dorodnitzyn', 'doubled', 'doubly', 'doubtful', 'downwards', 'dp', 'dramatic', 'draped', 'drastic', 'driest', 'drift', 'drilled', 'drilling', 'drive', 'droop', 'dropoff', 'dropped', 'drosophila', 'drum', 'du', 'ducing', 'ducting', 'dufour', 'duplicated', 'duplicating', 'duralumin', 'dust', 'dvl', 'dx', 'easier', 'easiest', 'echo', 'econd', 'economics', 'economy', 'eddyviscosity', 'edelfelt', 'edgedisplacements', 'edition', 'edpm', 'educational', 'edward', 'efficacious', 'efflux', 'effusion', 'egger', 'eigenvectors', 'einbinder', 'ejector', 'elapsed', 'elapsing', 'electro', 'electrode', 'electroformed', 'electrohydrodynamic', 'electrostatically', 'elegant', 'elemental', 'elevation', 'elevator', 'elimination', 'elliott', 'ellipselike', 'elliptically', 'embraced', 'embracing', 'embryonic', 'emerges', 'emphasising', 'ence', 'encircling', 'encompass', 'encompassing', 'endeavour', 'ended', 'endorsed', 'endplate', 'endurance', 'engesser', 'engined', 'england', 'enhance', 'enhanced', 'eniac', 'enjoy', 'enjoys', 'enlarge', 'enlarges', 'enlightening', 'enormously', 'enquiry', 'enskog', 'ensuing', 'enthalpics', 'entirety', 'entitled', 'entrained', 'entrant', 'envisaged', 'epoxy', 'equate', 'equatorial', 'equicohesive', 'equilateral', 'equilibrate', 'equivalently', 'er', 'erature', 'erdmann', 'erfc', 'erickson', 'erosive', 'erratic', 'eter', 'etude', 'eucken', 'eulerian', 'evaluates', 'evaporate', 'evaporates', 'eventual', 'evolve', 'evolved', 'evvard', 'exaggeration', 'examinable', 'examines', 'excessively', 'exchanger', 'excites', 'excused', 'exerted', 'exerts', 'exhaustive', 'exiting', 'exotic', 'expand', 'expands', 'expectation', 'expects', 'expended', 'expenditure', 'experimenter', 'explaining', 'explains', 'exploit', 'exploited', 'exploration', 'explore', 'explored', 'exploring', 'explosive', 'exponentiallyficients', 'expose', 'expounded', 'extensible', 'extinction', 'extra', 'extremal', 'eye', 'facet', 'facilitates', 'facilitating', 'fage', 'fales', 'fastened', 'fat', 'fd', 'featuring', 'feb', 'fed', 'feldman', 'fewer', 'ff', 'fibrous', 'fictitious', 'fictitiously', 'filter', 'filtering', 'financial', 'fine', 'finely', 'finest', 'finish', 'finitedifference', 'fire', 'firm', 'firmer', 'firstorder', 'fission', 'fix', 'fixedpitch', 'fl', 'flanged', 'flapped', 'flash', 'flatness', 'flatplate', 'flex', 'fllayer', 'flocculent', 'flowmeter', 'flowmeters', 'fluctuates', 'fluiddynamic', 'fluidized', 'fluttered', 'focus', 'focused', 'focusing', 'fog', 'forcible', 'fore', 'forebodies', 'forecast',

'forecasting', 'forefront', 'foreknowledge', 'formulate', 'formulating', 'formulism', 'fort', 'forty', 'forwardacceleration', 'forwardflight', 'foundation', 'fourcomponent', 'fourthorder', 'fourwave', 'fralich', 'framed', 'france', 'fredholm', 'freeing', 'freely', 'frictionless', 'friedman', 'frontal', 'frozenflow', 'fruit', 'frustration', 'furnace', 'furnishes', 'furnishing', 'fused', 'fuselagesection', 'garabedian', 'gardless', 'garrick', 'gascous', 'gasdynamical', 'gathered', 'gauged', 'geckeler', 'geiger', 'genuine', 'geoffrey', 'geometrics', 'geophysical', 'gerard', 'gershuni', 'getting', 'gibbs', 'giese', 'glassy', 'globule', 'glycerin', 'goal', 'goodier', 'gorcum', 'gordon', 'gortler', 'gothic', 'grad', 'graham', 'gram', 'grand', 'granular', 'grashof', 'grasshof', 'gravalos', 'grazing', 'gridwork', 'grimminger', 'grober', 'groove', 'grossman', 'grouped', 'growing', 'grunow', 'gw', 'gyration', 'gyroscope', 'gyroscopic', 'haag', 'hadamard', 'halfangle', 'halfway', 'halstead', 'halved', 'hamel', 'hammerhead', 'hampered', 'handicap', 'handle', 'handley', 'hansen', 'hantzsche', 'happen', 'happened', 'hardware', 'harris', 'hash', 'hasimoto', 'hastening', 'haveg', 'havelock', 'hay', 'hazard', 'headed', 'heaslet', 'heavier', 'heaving', 'heimel', 'heimenz', 'helical', 'heliocentric', 'hemispheroidal', 'hencky', 'hero', 'herrmann', 'hick', 'hidalgo', 'highenergy', 'higherorder', 'highfrequency', 'highvelocity', 'himself', 'hindered', 'hingeaxis', 'hirschfelder', 'hirshfelder', 'hise', 'historically', 'historics', 'hit', 'hitherto', 'hitting', 'hoffman', 'hogging', 'holder', 'holding', 'holland', 'hollowcylinder', 'homann', 'home', 'homing', 'homogenous', 'hope', 'hopelessly', 'hopf', 'hoshizaki', 'hosokawa', 'hotter', 'housefly', 'housing', 'hover', 'hovercraft', 'howorth', 'hpyersonic', 'hromas', 'huang', 'humphreys', 'hundred', 'hurd', 'hybrid', 'hydraulic', 'hydraulically', 'hydroballistic', 'hydrocarbon', 'hypass', 'hyperliptic', 'hypersoule', 'hysol', 'hysteretic', 'ia', 'idealizing', 'identically', 'identifiable', 'identifying', 'identity', 'iferation', 'igc', 'ignite', 'ignited', 'ignorable', 'ignores', 'igy', 'illingworth', 'illustrious', 'im', 'imaginary', 'imai', 'imilar', 'immersed', 'immobile', 'immovable', 'impair', 'imparted', 'implement', 'implied', 'imply', 'impression', 'imprimerie', 'improper', 'inaccessible', 'inadmissable', 'inadvertently', 'inapplicable', 'inappropriate', 'inasmuch', 'inboard', 'incidental', 'inclusive', 'incoming', 'inconclusive', 'inconel', 'inconsistency', 'inconsistent', 'incorporating', 'incrementalnormal', 'incurring', 'indefinitely', 'indetermining', 'indirectly', 'indium', 'inducedpressure', 'inductance', 'industry', 'inefficient', 'inertial', 'inevitably', 'inexpensive', 'inferred', 'infiniteaspect', 'influencing', 'influential', 'ing', 'inhomogeneous', 'inject', 'injectioncooled', 'injector', 'ink', 'inletmass', 'inoperative', 'inplane', 'inquire', 'inquired', 'insect', 'inserted', 'inserting', 'insertion', 'inspire', 'inst', 'instant', 'instigated', 'instruct', 'instruction', 'instrumented', 'insulates', 'insure', 'insuring', 'integrable', 'integrands', 'integrate', 'integrity', 'integrodifferential', 'intensification', 'intensified', 'intention', 'intentionally', 'interacted', 'interanl', 'interchangeable', 'intercrystalline', 'interdependence', 'interferograms', 'intergranular', 'interim', 'internat', 'interposed', 'interpretive', 'intersects', 'interstellar', 'intimately', 'intractable', 'intrinsically', 'introduring', 'intuition', 'invalidate', 'invalidated', 'invariably', 'inverts', 'inviscidflow', 'invoked', 'inwards', 'iodine', 'iowa', 'irbm', 'iron', 'irritant', 'isochronousstress', 'isofoam', 'isotope', 'isovel', 'item', 'jacobian', 'jan', 'janzen', 'japan', 'jeffery', 'jetboundary', 'jetexit', 'jetlike', 'jetstream', 'jmin', 'jmjn', 'johnson', 'join', 'jointly', 'jouguet', 'joule', 'judiciously', 'july', 'justifiable', 'justify', 'kallmann', 'katzen', 'kaye', 'keeping', 'keller', 'kelley', 'kemp', 'kempf', 'kempner', 'kendrick', 'kernal', 'kernelfunction', 'keulegan', 'keune', 'kg', 'kilometre', 'kinematically', 'kinked', 'kinking', 'kirk', 'klein', 'knee', 'knowing', 'ko', 'korst', 'krasilshchikova', 'krypton', 'kt', 'kuchemann', 'kuo', 'kussner', 'la', 'labour', 'lacking', 'lacquer', 'laguerre', 'laid', 'laminarboundary', 'laminarturbulent', 'laminate', 'land', 'langer', 'lap', 'laplacian', 'largeangled', 'largeaspect', 'largechord', 'largescale', 'laszlo', 'late', 'latent', 'laufer', 'launch', 'launching', 'layered', 'ld', 'leaf', 'lecture', 'leggett', 'legitimate', 'lender', 'lengthened', 'lennard', 'letter', 'letting', 'lever', 'liable', 'liapunov', 'libby', 'libration', 'librove', 'license', 'liftingsurface', 'lightly', 'lin', 'linder', 'lindsey', 'lineal', 'linearizedtheory', 'linearizing', 'liouville', 'lip', 'liquidinjection', 'listing', 'literal', 'lmethod', 'lobe', 'lobed', 'localize', 'localsimilarity', 'locate', 'locating', 'lockheed', 'locus', 'logically', 'loitsianskii', 'lomax', 'london', 'longest', 'longitudinalvortices', 'longtime', 'looking', 'loop', 'lose', 'lost', 'lotz', 'loudspeaker', 'louse', 'lowaspect', 'lowsubsonic', 'lowturbulence', 'ltd', 'lubricant', 'lucid', 'ludweig', 'ludwieg', 'lyon', 'lyskov', 'macadam', 'machined', 'machnumber', 'maeder', 'mag', 'magnesium', 'magnetoacoustic', 'magnetofluidmechanics', 'magnetogasdynamic', 'magnetoplasma', 'mahony', 'maintainance', 'maintaining', 'maintenance', 'majorminor', 'malfunction', 'maneuvering', 'manocuvres', 'manometer', 'manufacture', 'manufacturing', 'marin', 'maritime', 'mark', 'martin', 'mask', 'masked', 'maskell', 'maslen', 'massachusetts', 'massbalance', 'massbalanced', 'mat', 'matric', 'matting', 'maximuin', 'maxwell', 'mc', 'mccune', 'mckinnon', 'meansquare', 'meant', 'measurable', 'mech', 'mechs', 'meh', 'mehtod', 'meissner', 'meksyn', 'melamine', 'memorial', 'mentioning', 'merely', 'merger', 'merging', 'mesh', 'meteorite', 'methanol', 'mev', 'mhd', 'michael', 'michielsen', 'microinch', 'microinches', 'micromanometer', 'micron', 'microwave', 'midchord', 'midspan', 'midway', 'migrate', 'mil', 'mild', 'millikan', 'millimeter', 'mind', 'minimisation', 'minimised', 'minimises',

'minimizes', 'minnesota', 'mirels', 'mirror', 'misalinement', 'mismatch', 'mistake', 'mitigated', 'mixedflow', 'mm', 'modeled', 'modeling', 'modifies', 'modulating', 'mol', 'molten', 'molybdenum', 'momentary', 'monatomic', 'monograph', 'monopole', 'monoxide', 'morning', 'mostly', 'motionless', 'mound', 'mountain', 'mounting', 'mph', 'msvo', 'mth', 'muffler', 'muller', 'multiaxial', 'multibay', 'multicellular', 'multicomponent', 'multilayer', 'multiplied', 'multiplying', 'multistaging', 'mum', 'munk', 'murphy', 'murray', 'myself', 'nacelleexit', 'name', 'named', 'nationale', 'nautical', 'nd', 'neared', 'nearer', 'nearing', 'neater', 'necessitate', 'necessitates', 'negation', 'neighbouring', 'neumann', 'neutrally', 'newly', 'news', 'nice', 'nicolson', 'nielsen', 'nimonic', 'ninefold', 'nineteen', 'ninth', 'nitric', 'nocturnal', 'nomenclature', 'nominally', 'nomograph', 'nonablating', 'nonaligned', 'nonaxisymmetrical', 'noncirculatory', 'noncommunicating', 'nonconcurrence', 'nonconduction', 'noncorrosive', 'nondecomposing', 'nondimensionalized', 'nonexistence', 'nonflammable', 'nongyroscopic', 'nonheat', 'nonhomogeneous', 'noninclined', 'noninjection', 'noninsulated', 'nonisentropic', 'nonisothermal', 'nonnegative', 'nonorthogonal', 'nonpermeable', 'nonreacting', 'nonreactive', 'nonrigorous', 'nonsingular', 'nonslationary', 'nonslender', 'nonsphere', 'nonspherical', 'nonstationary', 'nontoxic', 'nontrivial', 'nonunity', 'normalpressure', 'norman', 'notable', 'noteworthy', 'noticeably', 'notion', 'nozzleexit', 'np', 'nullify', 'numberical', 'numger', 'ob', 'obeys', 'objectionable', 'obliquely', 'obliterated', 'obscure', 'observables', 'observational', 'observe', 'observes', 'observing', 'obtainec', 'obtainment', 'obviously', 'oc', 'occasionally', 'occupying', 'occuring', 'octagonal', 'octahedral', 'octave', 'ofattack', 'offset', 'ofr', 'ogival', 'ogivecylinders', 'ohm', 'old', 'onera', 'onetenth', 'onto', 'op', 'opaque', 'opened', 'openended', 'operates', 'operative', 'opportunely', 'opportunity', 'opposing', 'opposition', 'optic', 'optimal', 'ordered', 'ordering', 'orderof', 'ordinated', 'ordnance', 'organization', 'orgin', 'oriented', 'originate', 'originting', 'orlfice', 'orr', 'orthogonality', 'orthogonally', 'oscillates', 'oscillator', 'oscillograph', 'osoen', 'ostrach', 'ours', 'outermost', 'outwards', 'outweight', 'ovaluated', 'oven', 'overcomes', 'overestimation', 'overexpands', 'overexpansion', 'overhang', 'overlapping', 'overlook', 'overly', 'overriding', 'overshadow', 'oversimplification', 'oversimplified', 'overstability', 'overstatement', 'overswing', 'owen', 'oxide', 'packard', 'painleve', 'paint', 'palte', 'papreg', 'para', 'parabolae', 'parasol', 'parietale', 'paris', 'parkes', 'particledensity', 'partition', 'passenger', 'peclet', 'peculiar', 'peng', 'people', 'percentthick', 'perfectgas', 'perforation', 'perform', 'periments', 'periodically', 'permeability', 'permissable', 'perplexing', 'persh', 'persist', 'persisting', 'persists', 'personal', 'perturbing', 'pessimistic', 'ph', 'phasing', 'philos', 'phosphorescent', 'photoelastic', 'photomultipliers', 'phugoid', 'physcal', 'physiological', 'picard', 'pictured', 'piercy', 'pimple', 'pindroh', 'pinkerton', 'pinned', 'placement', 'planck', 'planetocentric', 'plastically', 'plateau', 'platinum', 'plausibly', 'playing', 'plied', 'pluming', 'plunging', 'pneumatic', 'pod', 'poetic', 'poincare', 'poinds', 'pointednose', 'pointing', 'poiscuille', 'polars', 'pole', 'polyaxial', 'polycrystalline', 'polygonal', 'polymer', 'polystyrene', 'polytechnic', 'ponderomotive', 'popular', 'popularly', 'portious', 'posed', 'possessing', 'postulate', 'postulating', 'potentialflow', 'potentially', 'potentiometer', 'potter', 'poudres', 'powering', 'powerplants', 'practicably', 'practitioner', 'prager', 'prebuckled', 'preceded', 'precession', 'precipitated', 'precipitously', 'precise', 'precisely', 'preclude', 'precludes', 'predictable', 'predictive', 'predigested', 'predominating', 'preheating', 'premise', 'premixed', 'prepare', 'preponderance', 'prescribing', 'preserved', 'press', 'pressing', 'pressureinteraction', 'pressureratio', 'prevail', 'prevailing', 'prevalence', 'prevented', 'preventing', 'prevents', 'previouslygiven', 'prewhirl', 'price', 'prikl', 'prime', 'prism', 'prismatic', 'proceeding', 'processed', 'producer', 'profound', 'programing', 'programming', 'projecting', 'prolate', 'prominence', 'promised', 'promoting', 'propagates', 'propellerresearch', 'proportionality', 'proportionate', 'proportioned', 'proportioning', 'proposal', 'proposel', 'proposition', 'propounded', 'protect', 'protecting', 'protects', 'protein', 'proton', 'protrude', 'protruding', 'protrusion', 'protuberance', 'provision', 'provisional', 'provoking', 'pseudo', 'psi', 'psig', 'psychological', 'publish', 'pursued', 'puzzled', 'pv', 'pyramidal', 'pyrex', 'quadratic', 'quadratically', 'quadruplet', 'quantum', 'quarterchord', 'quartic', 'quenches', 'quest', 'questionable', 'quickacting', 'quickly', 'quieting', 'quotation', 'quote', 'rabotnov', 'radartracking', 'radialband', 'radian', 'radiates', 'radome', 'radon', 'raise', 'rake', 'raleigh', 'rameter', 'ramp', 'rand', 'randomly', 'randomness', 'rapidity', 'ratiobeen', 'reactor', 'ready', 'reaffirmed', 'realgas', 'realised', 'realistically', 'realization', 'realize', 'reanalysis', 'reanalyze', 'rearranged', 'reattached', 'reattaches', 'reattaching', 'receives', 'reciprocally', 'reciprocity', 'recirculating', 'recognised', 'recognizing', 'recoverable', 'recovered', 'rectangle', 'rectilinear', 'recur', 'recurrence', 'recursion', 'redefinition', 'redirecting', 'reducedfrequency', 'reducedmodulus', 'reenters', 'reentrytype', 'reexamined', 'referenceenthalpy', 'refine', 'reflecting', 'reflex', 'reformulated', 'refracts', 'register', 'regularity', 'reichardt', 'reimpinge', 'reinforcement', 'reinvestigated', 'reiss', 'relativistic', 'relax', 'relaxed', 'relay', 'releasing', 'relevance', 'relieve', 'rely', 'remember', 'removable', 'removal', 'remove', 'render', 'repair', 'repeat', 'repercussion', 'repetition', 'replace', 'replacement', 'replaces', 'representable', 'reproducible', 'requency', 'request', 'requested', 'researcher', 'researchtype',

'researchvehicle', 'resembled', 'resembles', 'reservation', 'reshotko', 'resiliant', 'resist', 'resisted', 'resolve', 'resonable', 'resource', 'restores', 'restoring', 'restricts', 'resultantforce', 'retain', 'retaining', 'retrorocket', 'retrovelocity', 'returned', 'returning', 'reuss', 'revealing', 'reversibility', 'reversible', 'revert', 'reverts', 'revised', 'rewrite', 'rewriting', 'rf', 'rheological', 'rib', 'ribbon', 'riccati', 'richmond', 'ridge', 'rigidly', 'riolate', 'rising', 'rivet', 'rl', 'roar', 'rocketon', 'rod', 'roe', 'rogers', 'roghness', 'rose', 'rotate', 'rotated', 'rott', 'roughened', 'roughening', 'roundness', 'routhhurwitz', 'routinized', 'roy', 'royal', 'rpm', 'rsults', 'rt', 'rudimentary', 'runge', 'russian', 'safe', 'safely', 'sail', 'sakurai', 'salmon', 'sandpaper', 'sanger', 'satelite', 'saunders', 'scaled', 'scanning', 'scarshaack', 'scattering', 'schematic', 'schoenherr', 'schuh', 'schultz', 'schwarz', 'sci', 'scoop', 'scruton', 'seal', 'search', 'searching', 'seasonal', 'secant', 'sectioned', 'secure', 'security', 'sedimentation', 'sedov', 'seeded', 'seeding', 'seeking', 'segmented', 'select', 'selective', 'selfpreserving', 'semiballistic', 'semicircle', 'semidiameter', 'semiinfinite', 'senior', 'sensible', 'separable', 'september', 'sequel', 'serbin', 'serenity', 'served', 'service', 'servo', 'sesired', 'seventhpower', 'severely', 'severly', 'sextic', 'sgn', 'shadow', 'shallower', 'shank', 'sharpedge', 'sharpnosed', 'shearstress', 'sheath', 'shedding', 'sheltered', 'shercliff', 'shesterikov', 'shielded', 'shielding', 'shifting', 'shockboundary', 'shockdetachment', 'shocked', 'shockexpansion', 'shockinduced', 'shockless', 'shockstandoff', 'shockwaves', 'shorten', 'shortens', 'shorter', 'shortest', 'shortlifetime', 'shortly', 'shrouded', 'shuffle', 'sibulkin', 'sideforce', 'sideslipping', 'sight', 'sign', 'signal', 'sigularity', 'silencing', 'silica', 'simmons', 'simplifies', 'simulates', 'singleand', 'singlepass', 'singlestage', 'singletype', 'singly', 'sir', 'sizeable', 'sized', 'sketchy', 'skew', 'skewed', 'skewedparabolic', 'skill', 'skipping', 'skirt', 'skirted', 'slewed', 'slide', 'slider', 'slighlty', 'slipping', 'sloped', 'sloshing', 'slower', 'slowing', 'slug', 'smallness', 'smell', 'smoothing', 'smoothness', 'snow', 'soaking', 'socalled', 'society', 'solidconstruction', 'solides', 'solidification', 'solves', 'sommerfeld', 'sophisticated', 'soret', 'soundsource', 'spacewise', 'spanning', 'spark', 'sparkschlieren', 'specialise', 'specialization', 'specifiic', 'spectacular', 'spectrograph', 'spence', 'sperical', 'spheric', 'sphericalatmosphere', 'spiked', 'spillage', 'spindle', 'spinner', 'spiralling', 'spirit', 'spite', 'splitter', 'spoken', 'spontaneously', 'spot', 'sprague', 'spray', 'spreiter', 'spurious', 'sputnik', 'sr', 'stabilize', 'staff', 'staggered', 'standardize', 'standoff', 'statical', 'statistically', 'steadystate', 'steamline', 'steepest', 'steepness', 'stefan', 'stepdown', 'stepped', 'stimulates', 'stimulus', 'stine', 'stipulated', 'stochastic', 'stoker', 'stonecypher', 'stony', 'stored', 'strange', 'stratagem', 'stratiform', 'streak', 'streamlining', 'streamtubes', 'strengthening', 'stretched', 'stripanalysis', 'stroboscope', 'stroboscopic', 'stronger', 'strouhal', 'struck', 'sturm', 'subarcs', 'subjective', 'subroutine', 'subsidiary', 'substantiates', 'substituted', 'subtending', 'succeeded', 'succeeding', 'sucked', 'suffer', 'suffered', 'suffix', 'sugar', 'suitability', 'summation', 'summed', 'summing', 'supercircular', 'superfast', 'superficial', 'superiority', 'superposed', 'superscript', 'superseded', 'suppose', 'supposition', 'sure', 'surely', 'surfacetemperature', 'surpassed', 'surprising', 'surrounded', 'surveyed', 'sustains', 'sutton', 'swallow', 'sweeping', 'sweptwing', 'switching', 'symposium', 'synge', 'synthesis', 'synthesized', 'systematized', 'tabulate', 'tacitly', 'tailing', 'tailless', 'tailoring', 'tailwind', 'talbot', 'tan', 'tang', 'tao', 'task', 'taxiing', 'technologically', 'teeth', 'telemeter', 'telemetered', 'telemetering', 'teller', 'temprature', 'termed', 'termination', 'terminology', 'termperature', 'termwise', 'terrestrial', 'tetrachloride', 'themselves', 'theolar', 'theoretic', 'therefrom', 'thermalfatigue', 'thermoaeroelastic', 'thermoelastic', 'thermomechanical', 'thermomechanics', 'thermopropulsive', 'thesis', 'theta', 'thicken', 'thickened', 'thing', 'think', 'thins', 'thinwalled', 'thirdly', 'thn', 'thomas', 'thorough', 'threaten', 'threedimensionality', 'threshold', 'throttle', 'throw', 'tical', 'till', 'ting', 'tion', 'tional', 'tions', 'titan', 'tne', 'tnis', 'toambient', 'tobody', 'today', 'toke', 'told', 'tolerable', 'tollmienschlichting', 'tomomentum', 'tomotika', 'took', 'torda', 'toriconic', 'toriconical', 'totalangle', 'touched', 'towall', 'townend', 'townsend', 'toxic', 'tr', 'tracing', 'tracking', 'tractable', 'tractor', 'traditional', 'traight', 'trans', 'transcendant', 'transcendental', 'transducer', 'transferring', 'transfn', 'transit', 'translated', 'translatory', 'transpired', 'transportation', 'transversality', 'travel', 'traveled', 'traversal', 'traversed', 'traversing', 'trefftz', 'trefitz', 'tribute', 'trick', 'tried', 'triggering', 'trigonometrical', 'trochoidal', 'truncation', 'tsion', 'tubing', 'tubular', 'turbances', 'turbojetpowered', 'turbomachines', 'turned', 'twoand', 'twomode', 'typically', 'typifies', 'typographical', 'typography', 'uhf', 'ultra', 'ultracentrifuge', 'umerical', 'unablated', 'unaccompanied', 'unanticipated', 'unavoidable', 'unbalanced', 'unbanked', 'unblunted', 'uncertain', 'unconstrained', 'uncontrollable', 'uncorrected', 'undamped', 'underexpanded', 'underexpansion', 'undergoes', 'undergone', 'underneath', 'underside', 'understand', 'undersurface', 'undertook', 'undesirable', 'undetermined', 'undoubtedly', 'undrooped', 'unduly', 'unevenness', 'unexpected', 'unexpectedly', 'unexplained', 'unfamiliar', 'unfavourable', 'unfilled', 'unflanged', 'unfolding', 'unhappy', 'uni', 'unidirectional', 'unidirectionally', 'unification', 'unimportant', 'unintentionally', 'unitary', 'united', 'univeral', 'unmanned', 'unorthodox', 'unperturbed', 'unpowered', 'unpublished', 'unrealistic', 'unreality', 'unrelativistic', 'unresolved', 'unrestrained', 'unrestricted', 'unretarded', 'unseparated', 'unspecified',

'unsupported', 'unsymmetrical', 'untenable', 'untested', 'untrimmed', 'upflow', 'upward', 'urgent', 'usage', 'usefully', 'usefulness', 'user', 'utia', 'utilised', 'utilises', 'utilize', 'utilzation', 'ux', 'va', 'valuedness', 'valve', 'vanguard', 'vaporizing', 'vapour', 'variability', 'variabletemperature', 'variant', 'veiocity', 'velocitycomponent', 'velocitydistribution', 'velocityprofiles', 'venting', 'veocity', 'verfied', 'verifies', 'versatility', 'vertexcentered', 'verticaltake', 'vi', 'vibrated', 'viewing', 'vii', 'violates', 'violet', 'viscid', 'viscoelasticity', 'viscositytemperature', 'visibility', 'vision', 'visualised', 'visualising', 'visualized', 'vital', 'vlasov', 'volterra', 'volumetric', 'vz', 'waisted', 'wakelike', 'wanlass', 'warhead', 'warmer', 'warning', 'warped', 'washing', 'wassermann', 'wavefront', 'wavevelocity', 'waviness', 'wax', 'wb', 'weakened', 'weaker', 'weakly', 'weierstrass', 'weighs', 'weighted', 'weighting', 'weinstein', 'weitbrecht', 'wellestablished', 'wendt', 'wet', 'wheel', 'wherever', 'whichever', 'whirl', 'whisker', 'white', 'whitehead', 'widen', 'widespread', 'wiley', 'wilson', 'windstream', 'windtunnels', 'wingpropeller', 'wingtail', 'winny', 'wise', 'wissler', 'withdrawal', 'won', 'wooded', 'woodgate', 'workingsection', 'world', 'worsened', 'worthwhile', 'worthy', 'woven', 'wright', 'wrinkled', 'xenon', 'xx', 'yarymovych', 'yeh', 'yeilds', 'yuan', 'yy', 'zamm', 'zbrozek', 'zehnder', 'zerolift', 'zeroth', 'zhukhovitskii', 'zoom', 'zuk']

10. Dictionary Term with Largest and Lowest DFs in Index Version 2

VERSION 2 MAX and MIN DFs :

Max DF Terms(s) : ['to']

Min DF Terms(s) : ['ab', 'abbrevi', 'ablatedlength', 'abovement', 'absent', 'abundantli', 'academ', 'acceleromet', 'accentu', 'access', 'accident', 'accru', 'acet', 'acquisit', 'acr', 'acrodynam', 'acrothermochemistri', 'acrothermoelast', 'actur', 'adjacentequilibrium', 'adjoin', 'adjoint', 'administr', 'admixtur', 'adsorpt', 'advent', 'aeolotrop', 'aerfoil', 'aerial', 'aeroballist', 'aerocosmonaut', 'aerodynamieist', 'aerotherm', 'aerothermochem', 'aerothermodynam', 'aerothermoelast', 'affirm', 'afresh', 'afterflow', 'age', 'agenc', 'ailcron', 'airborn', 'aircrafttyp', 'airforc', 'airfram', 'airlin', 'airload', 'airscrew', 'airspe', 'akin', 'align', 'allen', 'allmov', 'allround', 'allud', 'almen', 'altogeth', 'aluminium', 'amer', 'america', 'amongst', 'ampl', 'ampli', 'analytici', 'andland', 'anh', 'anhedr', 'anisotropi', 'annex', 'announc', 'annual', 'annul', 'anov', 'anti', 'antielast', 'anyway', 'ao', 'apoge', 'apparentmass', 'appraoch', 'apropri', 'ar', 'ardc', 'ari', 'arrest', 'arrow', 'ascrib', 'asid', 'asimplifi', 'ask', 'aspectratio', 'assert', 'assort', 'assur', 'astronaut', 'astrophys', 'atlarg', 'attaindd', 'atten', 'aural', 'auspic', 'avalu', 'avov', 'avro', 'ax', 'axialthrust', 'backup', 'backward', 'bagley', 'bakanov', 'balsa', 'ban', 'bang', 'bare', 'barrel', 'basetyp', 'basin', 'batchelor', 'batdorf', 'bc', 'beat', 'bed', 'bedford', 'beg', 'belief', 'bellevil', 'belltyp', 'belt', 'bendingtors', 'bent', 'ber', 'bergdolt', 'berger', 'bernoulli', 'berthelot', 'bertram', 'beskin', 'beta', 'bevel', 'beyween', 'billow', 'bimetal', 'bimolecular', 'birnbaum', 'bisector', 'bleich', 'blend', 'blown', 'blowout', 'blue', 'bluish', 'blunttrail', 'board', 'boat', 'boiloff', 'bolshakov', 'bomb', 'bonu', 'boost', 'bore', 'bornel', 'boundwel', 'bracket', 'branchlin', 'braunschweig', 'breguet', 'bright', 'broke', 'brooklyn', 'brown', 'budianski', 'builtup', 'bulent', 'bulg', 'bulkhead', 'bullet', 'bump', 'buoyanc', 'buri', 'butler', 'buzz', 'bypass', 'cabin', 'calib', 'cambridg', 'camera', 'canard', 'cant', 'capillari', 'caplan', 'capsul', 'caravel', 'carborundum', 'caret', 'carrier', 'carter', 'cast', 'castigliano', 'castolit', 'cater', 'cathod', 'cellulos', 'centaur', 'centerof', 'centigrad', 'centrelin', 'centripet', 'centroid', 'certainti', 'cf', 'chain', 'chap', 'chapter', 'characteristi', 'charter', 'chase', 'chem', 'chemisorpt', 'chernyi', 'cheski', 'chester', 'chien', 'child', 'china', 'choleski', 'chu', 'circuit', 'circuitri', 'circulatori', 'circumlunar', 'clamshel', 'classifi', 'classificaiton', 'claus', 'clay', 'clearanc', 'climin', 'closest', 'closur', 'coalesc', 'coars', 'coast', 'coat', 'coefstant', 'coher', 'colder', 'colloc', 'colour', 'columnmateri', 'combustionh', 'combut', 'command', 'commenc', 'commun', 'compani', 'companion', 'compet', 'competit', 'complimentari', 'compound', 'compressibleflow', 'compression', 'conceptu', 'concoct', 'concret', 'condamp', 'conductioncontrol', 'conductionexpans', 'conductor', 'confid', 'conformaltransform', 'conidt', 'connexion', 'consier', 'consolid', 'const', 'constanc', 'constantdens', 'constantli', 'consumpt', 'contempl', 'context', 'contradict', 'convergentdiverg', 'convict', 'convolut', 'cooler', 'cope', 'copper', 'cornel', 'corollari', 'corpuscular', 'correlogram', 'cosin', 'cosmonaut', 'coulomb', 'council', 'count', 'counteract', 'counterclockwis', 'coupon', 'cp', 'crabtre', 'craft', 'crank', 'creas', 'creepbuckl', 'crest', 'crew', 'crinolin', 'crook', 'crosscorrel', 'crush', 'cumbersom', 'cure', 'curl', 'curtain', 'cuss', 'cutoff', 'cx', 'cylinderhemispher', 'daley', 'dalton', 'dampomet', 'dana', 'dashin', 'dashn', 'dashth',

'dashthat', 'daytim', 'debat', 'decad', 'decidedli', 'declassifi', 'declin', 'decompos', 'decomposit', 'deeper',
'deepest', 'deffer', 'defi', 'deflagr', 'degrad', 'degreec', 'degreer', 'deioniz', 'deisgn', 'delft', 'delimit', 'delin',
'demarc', 'denc', 'denomin', 'denser', 'dent', 'depict', 'deplet', 'depress', 'deryagin', 'desk', 'determinant', 'detra',
'devreas', 'dewpoint', 'diagon', 'diam', 'dicsov', 'dictat', 'dielectr', 'dilat', 'dilut', 'dipol', 'dirac', 'director',
'dirichlet', 'discretefrequ', 'disequilibrium', 'disloc', 'dismiss', 'dispar', 'dispers', 'dispos', 'dissert', 'dissimilar',
'dissociationrecombin', 'distant', 'diurnal', 'divers', 'diversifi', 'doctor', 'dodecagon', 'dome', 'doppler',
'dorodnitsyn', 'dorodnitzinhowarth', 'dorodnitzyn', 'doubli', 'dp', 'dramat', 'drape', 'driest', 'drift', 'dropoff',
'drosophila', 'drum', 'du', 'duce', 'dufour', 'duralumin', 'dust', 'dvl', 'dx', 'easier', 'easiest', 'echo', 'econd',
'economi', 'eddyviscos', 'edelfelt', 'edgedisplac', 'edit', 'edpm', 'educ', 'edward', 'efficaci', 'efflux', 'effus',
'egger', 'eigenvector', 'einbind', 'ejector', 'electro', 'electrod', 'electroform', 'electrohydrodynam', 'electrostat',
'eleg', 'elliott', 'ellipselik', 'embryon', 'emerg', 'emphasis', 'enc', 'encircl', 'endeavour', 'endors', 'endplat', 'endur',
'engess', 'england', 'eniac', 'enjoy', 'enlighten', 'enorm', 'enquiri', 'enskog', 'ensu', 'enthalp', 'entireti', 'entitl',
'entrant', 'envisag', 'epoxi', 'equatori', 'equicohes', 'equilater', 'equilibr', 'equilibria', 'er', 'eratur', 'erdmann',
'erfc', 'erickson', 'errat', 'eter', 'etud', 'eucken', 'eulerian', 'evvard', 'exagger', 'excus', 'exot', 'expend',
'expenditur', 'exponentiallyfici', 'expound', 'extinct', 'extra', 'eye', 'facet', 'fage', 'fale', 'fasten', 'fat', 'fd', 'feb',
'fed', 'feldman', 'fewer', 'ff', 'fibrou', 'financi', 'finest', 'finish', 'finitediffer', 'firm', 'firmer', 'firstord', 'fission',
'fixedpitch', 'fl', 'flash', 'flatplat', 'flex', 'fllayer', 'floccul', 'fluiddynam', 'fluidiz', 'focu', 'fog', 'forcibl', 'fore',
'forefront', 'foreknowledg', 'fort', 'forti', 'forwardacceler', 'forwardflight', 'foundat', 'fourcompon', 'fourthord',
'fourwav', 'fralich', 'franc', 'fredholm', 'freeli', 'frictionless', 'friedman', 'frontal', 'frozenflow', 'frustrat', 'furnac',
'fuse', 'fuselagesect', 'garabedian', 'gardless', 'garrick', 'gascou', 'gather', 'geckel', 'geiger', 'genuin', 'geoffrey',
'geophys', 'gerard', 'gershuni', 'gibb', 'gies', 'glassi', 'globul', 'glycerin', 'goal', 'goe', 'goodier', 'gorcum',
'gordon', 'gortler', 'gothic', 'grad', 'graham', 'gram', 'grand', 'granular', 'grashof', 'grasshof', 'gravalo', 'graze',
'gridwork', 'grimming', 'grober', 'groov', 'grossman', 'grunow', 'gw', 'gyrat', 'gyroscop', 'haag', 'hadamard',
'halfangl', 'halfway', 'halstead', 'hamel', 'hammerhead', 'hamper', 'handicap', 'handley', 'hansen', 'hantzsch',
'hardwar', 'harri', 'hash', 'hasimoto', 'hasten', 'haveg', 'havelock', 'hazard', 'heaslet', 'heav', 'heavier', 'heimel',
'heimenz', 'helic', 'heliocentr', 'hemispheroid', 'hencki', 'hero', 'herrmann', 'hick', 'hidalgo', 'highenergi',
'higherord', 'highfrequ', 'highveloc', 'himself', 'hinder', 'hingeaxi', 'hirschfeld', 'hirshfeld', 'hise', 'hitherto',
'hoffman', 'hog', 'holder', 'holland', 'hollowcylind', 'homann', 'hopelessli', 'hopf', 'hoshizaki', 'hosokawa',
'hotter', 'hous', 'housefli', 'hovercraft', 'howorth', 'hpyerson', 'hroma', 'huang', 'humphrey', 'hundr', 'hurd',
'hybrid', 'hydroballist', 'hydrocarbon', 'hypass', 'hyperlipt', 'hypersoul', 'hysol', 'hysteret', 'ia', 'ifer', 'igc', 'igi',
'illingworth', 'illustri', 'im', 'imaginari', 'imai', 'imilar', 'immers', 'immobil', 'immov', 'impair', 'impart',
'implement', 'impress', 'imprimeri', 'improp', 'inaccess', 'inadmiss', 'inadvert', 'inapplic', 'inappropri',
'inasmuch', 'inboard', 'incident', 'incom', 'inconclus', 'inconel', 'incrementalnorm', 'indefinit', 'indetermin',
'indirectli', 'indium', 'inducedpressur', 'industri', 'ineffici', 'inerti', 'inexpens', 'infiniteaspect', 'influenti', 'ing',
'inhomogen', 'injectioncool', 'injector', 'ink', 'inletmass', 'inop', 'inplan', 'insect', 'inspir', 'inst', 'instant', 'instig',
'integrand', 'integrodifferenti', 'intensif', 'intensifi', 'interanl', 'interchang', 'intercrystallin', 'interdepend',
'interferogram', 'intergranular', 'interim', 'internat', 'interpos', 'interstellar', 'intract', 'introdur', 'intuit',
'inviscidflow', 'invok', 'iodin', 'iowa', 'irbm', 'iron', 'irrit', 'isochronousstress', 'isofoam', 'isotop', 'item',
'jacobian', 'jan', 'janzen', 'japan', 'jefferi', 'jetboundari', 'jetexit', 'jetlik', 'jetstream', 'jmin', 'jmjn', 'johnson',
'jointli', 'jouguet', 'joul', 'judici', 'juli', 'kallmann', 'katzen', 'kay', 'keller', 'kelley', 'kemp', 'kempf', 'kempner',
'kendrick', 'kernal', 'kernelfunct', 'keulegan', 'keun', 'kg', 'kilometr', 'kirk', 'klein', 'knee', 'ko', 'korst',
'krasilshchikova', 'krypton', 'kt', 'kuchemann', 'kuo', 'kussner', 'la', 'labour', 'lacquer', 'laguerr', 'laid', 'lamin',
'laminarboundari', 'laminarturbul', 'langer', 'lap', 'laplacian', 'largeangl', 'largeaspect', 'largechord', 'largescal',
'laszlo', 'late', 'latent', 'laufer', 'ld', 'lectur', 'leggett', 'legitim', 'lender', 'lengthen', 'lennard', 'letter', 'lever', 'liabl',
'liapunov', 'libbi', 'librat', 'librov', 'lice', 'licens', 'liftingsurfac', 'lightli', 'lin', 'linder', 'lindsey', 'lineal',
'linearizedtheori', 'liouvil', 'lip', 'liquidinject', 'lmethod', 'localsimilar', 'lockhe', 'locu', 'loitsianskii', 'lomax',
'london', 'longest', 'longitudinalvortic', 'longtim', 'look', 'loop', 'lose', 'lost', 'lotz', 'loudspeak', 'lowaspect',
'lowsubson', 'lowturbul', 'ltd', 'lucid', 'ludweig', 'ludwieg', 'lyon', 'lyskov', 'macadam', 'machnumb', 'maeder',
'mag', 'magnesium', 'magnetoacoust', 'magnetofluidmechan', 'magnetogasdynam', 'magnetoplasma', 'mahoni',
'mainten', 'majorminor', 'malfunct', 'manocuvr', 'manomet', 'marin', 'maritim', 'martin', 'maskel', 'maslen',
'massachusett', 'massbal', 'massbalanc', 'maximuin', 'maxwel', 'mc', 'mccune', 'mckinnon', 'meansquar', 'meant',
'meh', 'mehtod', 'meissner', 'meksyn', 'melamin', 'memori', 'merger', 'mesh', 'meteorit', 'methanol', 'mev', 'mhd',
'michael', 'michielsen', 'micromanomet', 'micron', 'microwav', 'midchord', 'midspan', 'midway', 'migrat', 'mil',
'mild', 'millikan', 'millimet', 'mind', 'minnesota', 'mirel', 'mirror', 'misalin', 'mismatch', 'mistak', 'mitig',

'mixedflow', 'mm', 'mol', 'molten', 'molybdenum', 'momentari', 'monatom', 'monograph', 'monopol', 'monoxid', 'morn', 'mostli', 'motionless', 'mound', 'mountain', 'mph', 'ms', 'msvo', 'mth', 'muffler', 'muller', 'multiaxi', 'multibay', 'multicellular', 'multicompon', 'multilay', 'mum', 'munk', 'murphi', 'murray', 'myself', 'nacelleexit', 'national', 'nautic', 'nd', 'nearer', 'neater', 'negat', 'neighbour', 'neumann', 'newli', 'news', 'nice', 'nicolson', 'nielsen', 'nimon', 'ninefold', 'nineteen', 'ninth', 'nitric', 'nocturn', 'nomenclatur', 'nomograph', 'nonabl', 'nonalign', 'noncirculatori', 'noncommun', 'nonconcurr', 'noncorros', 'nondecompos', 'nonexist', 'nonflamm', 'nongyroscop', 'nonheat', 'nonhomogen', 'noninclin', 'noninject', 'noninsul', 'nonisentrop', 'nonisotherm', 'nonneg', 'nonorthogon', 'nonperm', 'nonrigor', 'nonsingular', 'nonslationari', 'nonslend', 'nonstationari', 'nontox', 'nontrivi', 'nonun', 'normalpressur', 'norman', 'notabl', 'noteworthi', 'notion', 'nozzleexit', 'np', 'nullifi', 'numger', 'ob', 'objection', 'obliter', 'obscur', 'obtainec', 'obvious', 'oc', 'occasion', 'occupi', 'octagon', 'octahedr', 'octav', 'ofattack', 'offset', 'ofr', 'ogivecylind', 'ohm', 'old', 'onera', 'onetenth', 'onto', 'op', 'opaqu', 'openend', 'orderof', 'ordnanc', 'organ', 'orgin', 'origint', 'orlfic', 'orr', 'oscillograph', 'osoen', 'ostrach', 'our', 'outermost', 'outweight', 'ovalu', 'oven', 'overexpans', 'overhang', 'overli', 'overlook', 'overrid', 'overshadow', 'oversimplif', 'oversimplifi', 'overst', 'overstat', 'oversw', 'owen', 'packard', 'painlev', 'paint', 'palt', 'papreg', 'para', 'parabola', 'parasol', 'pari', 'parietal', 'park', 'particledens', 'partit', 'passeng', 'peclet', 'peng', 'peopl', 'percentthick', 'perfectga', 'periment', 'perplex', 'persh', 'person', 'pessimist', 'ph', 'philo', 'phosphoresc', 'photoelast', 'phugoid', 'physcal', 'physiolog', 'picard', 'pierci', 'pimpl', 'pin', 'pindroh', 'pinkerton', 'placement', 'planck', 'planetocentr', 'plateau', 'platinum', 'pli', 'plung', 'pneumat', 'pod', 'poetic', 'poincar', 'poind', 'pointednos', 'poiscuil', 'pole', 'polyaxi', 'polycrystallin', 'polym', 'polystyren', 'polytechn', 'ponderomot', 'popular', 'popularli', 'portiou', 'pose', 'potentialflow', 'potentiomet', 'potter', 'poudr', 'powerpl', 'practition', 'prager', 'prebuckl', 'precess', 'preclud', 'predigest', 'preheat', 'premis', 'premix', 'preponder', 'pressureinteract', 'pressureratio', 'preval', 'previouslygiven', 'prewhirl', 'price', 'prikl', 'prime', 'prism', 'prismat', 'profound', 'prolat', 'promin', 'propellerresearch', 'proposel', 'proposit', 'propound', 'protein', 'proton', 'protrus', 'protuber', 'provis', 'provision', 'provok', 'pseudo', 'psi', 'psig', 'psycholog', 'pursu', 'puzzl', 'pv', 'pyramid', 'pyrex', 'quadruplet', 'quantum', 'quarterchord', 'quartic', 'quest', 'quickact', 'quickli', 'quiet', 'quotat', 'rabotnov', 'radartrack', 'radialband', 'radian', 'radom', 'radon', 'rake', 'raleigh', 'ramet', 'ramp', 'rand', 'randomli', 'ratiobeen', 'reactor', 'readi', 'reaffirm', 'realga', 'realis', 'reanalysi', 'reanalyz', 'rearrang', 'recognis', 'recov', 'recover', 'rectangl', 'rectilinear', 'recur', 'recurr', 'recurs', 'redefinit', 'redirect', 'reducedfrequ', 'reducedmodulu', 'reenter', 'reentrytyp', 'referenceenthalpi', 'reflex', 'reformul', 'regist', 'reichardt', 'reimping', 'reinvestig', 'reiss', 'relativist', 'relay', 'reli', 'reliev', 'rememb', 'repair', 'repercuss', 'repetit', 'requenc', 'request', 'researchtyp', 'researchvehicl', 'reserv', 'reshotko', 'resili', 'resourc', 'resultantforc', 'retrorocket', 'retroveloc', 'reuss', 'rf', 'rheolog', 'rib', 'ribbon', 'riccati', 'richmond', 'ridg', 'rigidli', 'riolat', 'rivet', 'rl', 'roar', 'rocketon', 'rod', 'roe', 'roger', 'rogh', 'rose', 'rott', 'roughen', 'routhhurwitz', 'roy', 'royal', 'rpm', 'rsult', 'rt', 'rudimentari', 'rung', 'russian', 'sail', 'sakurai', 'salmon', 'sandpap', 'sanger', 'satelit', 'saunder', 'scan', 'scarshaack', 'schemat', 'schoenherr', 'schuh', 'schultz', 'schwarz', 'sci', 'scoop', 'scruton', 'secant', 'sediment', 'sedov', 'seed', 'selfpreserv', 'semiballist', 'semicircl', 'semidiamet', 'semiinfinit', 'senior', 'sensibl', 'septemb', 'sequel', 'serbin', 'seren', 'servic', 'servo', 'sesir', 'seventhpow', 'severli', 'sextic', 'sgn', 'shadow', 'shank', 'sharpedg', 'sharpnos', 'shearstress', 'sheath', 'shelter', 'shercliff', 'shesterikov', 'shockboundari', 'shockdetach', 'shockexpans', 'shockinduc', 'shockless', 'shockstandoff', 'shorter', 'shortest', 'shortli', 'shortlifetim', 'shuffl', 'sibulkin', 'sideforc', 'sight', 'sign', 'signal', 'sigular', 'silenc', 'silica', 'simmon', 'singleand', 'singlepass', 'singlestag', 'singletyp', 'singli', 'sir', 'sizeabl', 'sketchi', 'skewedparabol', 'skill', 'slew', 'slide', 'slider', 'slighlti', 'slosh', 'slower', 'slug', 'smell', 'snow', 'soak', 'socal', 'societi', 'solidconstruct', 'solidif', 'sommerfeld', 'sophist', 'soret', 'soundsourc', 'spacewis', 'spark', 'sparkschlieren', 'specialis', 'spectacular', 'spectrograph', 'spenc', 'speric', 'sphericalatmospher', 'spillag', 'spindl', 'spinner', 'spirit', 'spite', 'splitter', 'spoken', 'spontan', 'spot', 'spragu', 'spray', 'spreiter', 'spuriou', 'sputnik', 'sr', 'staff', 'stagger', 'standoff', 'steadyst', 'steamlin', 'steepest', 'stefan', 'stepdown', 'stimulu', 'stine', 'stipul', 'stochast', 'stoker', 'stonecyph', 'stoni', 'strang', 'stratagem', 'stratiform', 'streak', 'strengthen', 'stripanalysi', 'stronger', 'strouhal', 'struck', 'sturm', 'subarc', 'subroutin', 'subsidiari', 'subtend', 'suck', 'suffix', 'sugar', 'summat', 'supercircular', 'superfast', 'superfici', 'superpos', 'superscript', 'supersed', 'supposit', 'surfacetemperatur', 'surpass', 'surpris', 'sutton', 'swallow', 'sweptw', 'switch', 'symposium', 'syng', 'synthes', 'synthesi', 'tacitli', 'tailless', 'tailwind', 'talbot', 'tan', 'tang', 'tao', 'task', 'taxi', 'teeth', 'teller', 'tempratur', 'terminolog', 'termperatur', 'termwis', 'terrestri', 'tetrachlorid', 'themselv', 'theolar', 'therefrom', 'thermalfatigu', 'thermoaeroelast', 'thermoelast', 'thermomechan', 'thermopropuls', 'thesi', 'theta', 'thing', 'think', 'thinwal', 'thirdli', 'thn', 'thoma', 'thorough', 'threaten', 'threshold', 'throttl', 'throw', 'tical', 'till', 'ting', 'tional', 'titan', 'tne', 'tni', 'toambient', 'tobodi', 'today', 'toke', 'told', 'tollmienschlicht', 'tomomentum', 'tomotika', 'took', 'torda',

'toricon', 'totalangl', 'towal', 'townend', 'townsend', 'toxic', 'tr', 'tractabl', 'tractor', 'tradit', 'traight', 'tran', 'transcend', 'transcendent', 'transduc', 'transfn', 'translatori', 'trefftz', 'trefitz', 'tri', 'tribut', 'trick', 'trigger', 'trochoid', 'tsion', 'tubular', 'turbanc', 'turbojetpow', 'turbomachin', 'twoand', 'twomod', 'typograph', 'typographi', 'uhf', 'ultra', 'ultracentrifug', 'umer', 'unabl', 'unaccompani', 'unanticip', 'unavoid', 'unbalanc', 'unbank', 'unblunt', 'uncertain', 'unconstrain', 'uncontrol', 'uncorrect', 'undamp', 'underexpand', 'underexpans', 'undergon', 'underneath', 'undersid', 'undersurfac', 'undertook', 'undesir', 'undetermin', 'undoubtedli', 'undroop', 'unduli', 'uneven', 'unexpect', 'unexpectedli', 'unexplain', 'unfamiliar', 'unfavour', 'unfil', 'unflang', 'unfold', 'unhappi', 'uni', 'unidirect', 'unif', 'unimport', 'unintent', 'unitari', 'univer', 'unman', 'unorthodox', 'unperturb', 'unpow', 'unpublish', 'unreal', 'unrealist', 'unrelativist', 'unresolv', 'unrestrain', 'unrestrict', 'unretard', 'unsepar', 'unspecifi', 'unsupport', 'unsymmetr', 'unten', 'untest', 'untrim', 'upflow', 'upward', 'urgent', 'usag', 'user', 'utia', 'utilz', 'ux', 'va', 'valued', 'valv', 'vanguard', 'vapour', 'variabletemperatur', 'variant', 'veiociti', 'velocitycompon', 'velocitydistribut', 'velocityprofil', 'vent', 'veociti', 'verfi', 'versatil', 'vertexcent', 'verticaltak', 'vi', 'vii', 'violat', 'violet', 'viscid', 'viscositytemperatur', 'vision', 'vital', 'vlasov', 'volterra', 'volumetr', 'vs', 'vz', 'waist', 'wakelik', 'wanlass', 'warhead', 'warmer', 'warn', 'warp', 'wash', 'wassermann', 'wavefront', 'waveveloc', 'wax', 'wb', 'weaker', 'weakli', 'weierstrass', 'weigh', 'weinstein', 'weitbrecht', 'wellestablish', 'wendt', 'wheel', 'wherev', 'whichev', 'whisker', 'white', 'whitehead', 'widen', 'widespread', 'wiley', 'wilson', 'windstream', 'wingpropel', 'wingtail', 'winni', 'wise', 'wissler', 'withdraw', 'won', 'woodgat', 'workingsect', 'world', 'worsen', 'worthi', 'worthwhil', 'woven', 'wright', 'xenon', 'xx', 'yarymovych', 'yeh', 'yeild', 'yuan', 'yy', 'zamm', 'zbrozek', 'zehnder', 'zerolift', 'zeroth', 'zhukhovitskii', 'zoom', 'zuk']

11. Documents with largest doclen and maxtf in the collection:

Document with Max Doc Length : 1313 with doc len = 644

Document with largest MaxTF : 673 with Max TF = 26

Output Screenshots:

```
1. Time taken for building indices:
 1a. Time taken for building Index_Version1.uncompress : 5.973879337310791
 1b. Time taken for building Index_Version2.uncompress : 6.369318723678589
 1c. Time taken for building Index_Version1.compress   : 0.6033968925476074
 1d. Time taken for building Index_Version2.compress   : 0.5904693603515625
2. Size of Index_Version1.uncompress 1757148
3. Size of Index_Version2.uncompress 1685572
4. Size of Index_Version1.compress 2051103
5. Size of Index_Version2.compress 1743903
6. Number of Postings in each Version:
 6a. Number of Postings in Version1 Index: 6299
 6b. Number of Postings in Version2 Index: 4498
```

```
7. Postings for given list of words

TERM : reynolds
 Df : 192
 Total TF : 358
 Inverted Posting Lentgh (in Bytes) : 1800

TERM : nasa
 Df : 7
 Total TF : 6
 Inverted Posting Lentgh (in Bytes) : 160

TERM : prandtl
 Df : 63
 Total TF : 79
 Inverted Posting Lentgh (in Bytes) : 576

TERM : flow
 Df : 678
 Total TF : 1696
 Inverted Posting Lentgh (in Bytes) : 5888

TERM : pressure
 Df : 529
 Total TF : 1227
 Inverted Posting Lentgh (in Bytes) : 4576

TERM : boundary
 Df : 457
 Total TF : 997
 Inverted Posting Lentgh (in Bytes) : 4024

TERM : shock
 Df : 235
 Total TF : 658
 Inverted Posting Lentgh (in Bytes) : 2072
```

```
8. DF,TF and Postings for NASA
  DF for NASA : 7

   Entry 1 values :
    TF      :  1
    Doclen :  103
    MaxTF  :  4

   Entry 2 values :
    TF      :  1
    Doclen :  100
    MaxTF  :  4

   Entry 3 values :
    TF      :  1
    Doclen :  195
    MaxTF  :  6
```

```
11. Documents with largest doclen and maxtf in the collection

   Document with Max Doc Length : 1313 with doc len = 644
   Document with largest MaxTF  : 673 with Max TF = 26
```

Files Generated:

```
-rw------- 1 axb200027 cs  195404 Mar 31 14:07 lemma_posting_1
-rw------- 1 axb200027 cs  181821 Mar 31 14:07 stem_posting_1
-rw------- 1 axb200027 cs  197785 Mar 31 14:07 lemma_posting_2
-rw------- 1 axb200027 cs  184432 Mar 31 14:07 stem_posting_2
-rw------- 1 axb200027 cs  191169 Mar 31 14:07 lemma_posting_3
-rw------- 1 axb200027 cs  178237 Mar 31 14:07 stem_posting_3
-rw------- 1 axb200027 cs  192153 Mar 31 14:07 lemma_posting_4
-rw------- 1 axb200027 cs  178987 Mar 31 14:07 stem_posting_4
-rw------- 1 axb200027 cs  196199 Mar 31 14:07 lemma_posting_5
-rw------- 1 axb200027 cs  182219 Mar 31 14:07 stem_posting_5
-rw------- 1 axb200027 cs  193658 Mar 31 14:07 lemma_posting_6
-rw------- 1 axb200027 cs  180191 Mar 31 14:07 stem_posting_6
-rw------- 1 axb200027 cs  193350 Mar 31 14:07 lemma_posting_7
-rw------- 1 axb200027 cs  179612 Mar 31 14:07 stem_posting_7
-rw------- 1 axb200027 cs  195169 Mar 31 14:07 lemma_posting_8
-rw------- 1 axb200027 cs  182153 Mar 31 14:07 stem_posting_8
-rw------- 1 axb200027 cs  193095 Mar 31 14:07 lemma_posting_9
-rw------- 1 axb200027 cs  179696 Mar 31 14:07 stem_posting_9
-rw------- 1 axb200027 cs  193502 Mar 31 14:07 lemma_posting_10
-rw------- 1 axb200027 cs  179537 Mar 31 14:07 stem_posting_10
-rw------- 1 axb200027 cs   50222 Mar 31 14:07 lemma_posting_11
-rw------- 1 axb200027 cs   45968 Mar 31 14:07 stem_posting_11
-rw------- 1 axb200027 cs 1757148 Mar 31 14:07 Index_Version1.uncompress
-rw------- 1 axb200027 cs 1685572 Mar 31 14:07 Index_Version2.uncompress
-rw------- 1 axb200027 cs 2051103 Mar 31 14:07 Index_Version1.compress
-rw------- 1 axb200027 cs   90137 Mar 31 14:07 Dict_Version1.compress
-rw------- 1 axb200027 cs 1743903 Mar 31 14:07 Index_Version2.compress
-rw------- 1 axb200027 cs   64395 Mar 31 14:07 Dict_Version2.compress
```